Multiple Object Properties Drive Scene-Selective Regions

Vanessa Troiani, Anthony Stigliani, Mary E. Smith and Russell A. Epstein

Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania, Philadelphia, PA, USA

Address correspondence to Russell A. Epstein, 3720 Walnut St., Philadelphia, PA 19104. Email: epstein@psych.upenn.edu

Neuroimaging studies have identified brain regions that respond preferentially to specific stimulus categories, including 3 areas that activate maximally during viewing of real-world scenes: The parahippocampal place area (PPA), retrosplenial complex (RSC), and transverse occipital sulcus (TOS). Although these findings suggest the existence of regions specialized for scene processing, this interpretation is challenged by recent reports that activity in scene-preferring regions is modulated by properties of isolated single objects. To understand the mechanisms underlying these object-related responses, we collected functional magnetic resonance imaging data while subjects viewed objects rated along 7 dimensions, shown both in isolation and on a scenic background. Consistent with previous reports, we find that scene-preferring regions are sensitive to multiple object properties; however, results of an item analysis suggested just 2 independent factors—visual size and the landmark suitability of the objects—sufficed to explain most of the response. This object-based modulation was found in PPA and RSC irrespective of the presence or absence of a scenic background, but was only observed in TOS for isolated objects. We hypothesize that scene-preferring regions might process both visual qualities unique to scenes and spatial qualities that can appertain to either scenes or objects.

Keywords: fMRI, parahippocampal place area, scene perception, spatial layout, visual context

Introduction

Human visual cortex exhibits a high degree of functional organization. Most notably, neuroimaging studies have identified regions in the ventral visual stream that respond preferentially to certain stimulus categories, including faces (Kanwisher et al. 1997), bodies (Downing et al. 2001), and scenes (Epstein and Kanwisher 1998). Although the existence of these regionally specific stimulus preferences is not in question, the interpretation of these findings has been the topic of considerable debate. Some authors have focused on the maximal responses in these regions, arguing that they indicate the operation of specialized modules dedicated to recognition of ecologically significant stimulus classes (Kanwisher 2010). Others have focused on the submaximal responses within these regions, arguing that they indicate the operation of more general-purpose recognition mechanisms, whose engagement is not tied by necessity to any specific stimulus class (Gauthier et al. 1999; Bar 2004).

Here, we consider this matter for regions that respond primarily to visual scenes. By a "scene," we mean a section of a real-world environment (or an artificial equivalent) that includes both foreground objects and fixed background elements and can be ascertained in a single view (Henderson and Hollingsworth 1999; Epstein and MacEvoy 2011). For example, a photograph of a room, a landscape, or a city street

is a scene (or, more precisely, an image of a scene). We contrast scenes with "objects" such as automobiles, blenders, shoes, and screwdrivers, which are discrete entities bounded by a single contour that do not have background elements. Three regions of the human brain have been identified that are sensitive to this distinction insofar as they respond more strongly to scenes than to objects: The parahippocampal place area (PPA), the retrosplenial complex (RSC), and the transverse occipital sulcus (TOS).

Of these 3 regions, the PPA has been most strongly implicated in scene recognition, whereas the RSC has been more strongly implicated in spatial memory retrieval, and the function of TOS is unknown (Epstein et al. 2007; Park et al. 2007). Recent studies have demonstrated a close link between scene recognition and neural activity in the PPA by showing that multivoxel patterns in the PPA distinguish between scene categories (Walther et al. 2009, 2011; Epstein and Morgan 2011) and scene exemplars (Morgan et al. 2011) and are predictive of behavioral recognition performance (Walther et al. 2009). Although it is not entirely clear that stimulus features drive the neural distinctions between scene categories and scene exemplars, one possibility is that they relate to differences in the global geometric features of scenes, such as the degree of openness or closedness (Ross and Oliva 2010; Kravitz et al. 2011; Park et al. 2011). In any case, the evidence that the PPA is involved in scene processing—and, in particular, processing of spatial information associated with scenes—is quite strong.

However, the fact that the PPA—as well as the RSC and TOS—respond preferentially to scenes does not necessarily mean that these regions respond to scenes per se; rather, they might instead be responsive to a stimulus quality that could be shared by both scenes and some kinds of objects. In particular, it is possible that the PPA (or other scene-responsive regions) might process spatial information that could adhere to single, individual objects under the right circumstances. Relatedly, it is possible that response in these regions might depend, in part, on high-level interpretation of the stimulus: Some stand-alone objects might be treated as landmarks (i.e. fixed entities that act as a navigational reference) or signifiers of a place (i.e. specific location in the world).

Indeed, even the very first report on the PPA provided some evidence that it discriminates between nonscene categories, as the responses to houses and buildings (shown without background) was numerically greater than the response to common everyday objects, and the response to common everyday objects was numerically greater than the response to faces (Epstein and Kanwisher 1998). Although buildings might be considered partial scenes, even when they are shown removed from their surroundings, objects such as blenders, toasters, and chairs are clearly not scenes. Thus, the pattern of greater response to objects than to faces seems hard to reconcile with the idea that the PPA exclusively

processes scenes, or spatial properties that are unique to scenes. Furthermore, multivoxel patterns in the PPA have been shown to distinguish between stand-alone objects under some circumstances (Diana et al. 2008; MacEvoy and Epstein 2011), but not others (Spiridon and Kanwisher 2002). All these data suggest a possible role for the PPA in processing information obtainable from single objects.

Consistent with this view, a number of recent studies have reported that the PPA responds to the spatial qualities of objects. For example, Konkle and Oliva (2011) showed that a region of posterior parahippocampal cortex (pPHC) that partially overlaps with the PPA responds more strongly to large objects (e.g. car, piano) than to small objects (e.g. strawberry, calculator), even when the stimuli have equivalent retinal size. Similarly, Amit et al. (2012) and Cate et al. (2011) observed greater pPHC activity to large and distant objects than to small and nearby objects, where size and distance were implied by the presence of Ponzo lines defining a minimal scene. In a related vein, Mullally and Maguire (2011) reported greater pPHC activation when subjects imagined objects that conveyed a strong sense of surrounding space than when they imagined objects that had weak "spatial definition." Finally, Bar and Aminoff (2003); Bar (2004); and Bar et al. (2008) reported greater pPHC activity during viewing of objects with strong contextual associations than during viewing of objects with weak contextual associations (but see Epstein and Ward 2010). This context effect was found in pPHC for spatial contextual associations but not for equivalent nonspatial contextual associations (Aminoff et al. 2007). All of these characteristics—physical size, distance, spatial definition, and contextual associations—would be expected to be greater for scenes than for most objects and, thus, could potentially explain the previously described "categorical" response to scenes.

Here, we propose a reconceptualization of the function of the PPA that builds upon this previous work. Rather than considering the PPA as a scene area exclusively, or as an area that processes a single kind of object property, we propose that the region is sensitive to multiple physical and conceptual properties that relate to the suitability of the stimulus as a navigationally relevant item, in addition to being sensitive to properties that are unique to scenes. Thus, in contrast to the aforementioned studies, which tend to focus on a single spatial property in driving a PPA response, we propose that there are several physical and conceptual properties that might act as suitable cues to indicate that a stimulus is a "landmark" or a "place." To test this idea, we examined the contribution of 7 stimulus-based properties on PPA activity. These included 6 object-based properties, which were assessed by independent subject ratings (placeness, spatial definition, contextual associations, fixedness, distance, and physical size), and also the visual size of the object. We also examined the effect of presenting objects both with and without the scenic background. To anticipate, our results suggest that the 6 object-based properties are effective but redundant drivers of response in the PPA and also in the RSC.

Materials and Methods

Subjects

Twelve right-handed subjects (4 males; ages 19–27 years) with normal or corrected-to-normal vision were recruited from the University of Pennsylvania community to participate in the main functional magnetic resonance imaging (fMRI) experiment. Other subjects meeting the same criteria rated the stimuli in behavioral studies and online surveys. Specifically, 86 subjects rated object stimuli in laboratory sessions (placeness, $n = 15$; spatial definition, $n = 15$; contextual associations, $n = 15$; fixedness, $n = 20$; physical size, $n = 21$), while 24 subjects rated distance online. All participants gave written informed consent in accordance with procedures approved by the University of Pennsylvania Institutional Review Board and were paid for their participation.

Magnetic Resonance Imaging Acquisition

Scans were performed at the Hospital of the University of Pennsylvania on a 3-T Siemens Trio scanner equipped with a Siemens body coil and an 8-channel multiple-array Nova Medical head coil. Structural T_1-weighted images for anatomical localization were acquired using 3-dimensional MPRAGE pulse sequences (time repetition [TR] = 1650 ms, time echo [TE] = 3 ms, time to inversion = 950 ms, voxel size = 0.9766 × 0.9766 × 1.0 mm, matrix size = 192 × 256 × 160). T_2-weighted scans sensitive to blood oxygenation level-dependent (BOLD) contrasts were acquired using a gradient-echo, echo-planar pulse sequence (TR = 3000 ms, TE = 30 ms, voxel size = 3 × 3 × 3 mm, matrix size = 64 × 64 × 45). Stimuli were displayed on a rear projected mylar screen using an Epson 8100 3-LCD projector and viewed through a mirror mounted to the head coil, and responses were recorded using a 4-button fiber-optic response pad device.

Stimuli

Stimuli were 200 768 × 768 pixel color scene images depicting a single focal object with surrounding background elements (including other objects) and 200 altered versions of these scenes in which the focal object was shown in isolation in the same screen position with the background elements removed. Thus, there were 2 versions of each stimulus: A Scene version (object + background) and an Object version (object + no background). Each subject in the fMRI experiment saw every focal object once, in either the Scene or Object configuration; choice of version was counterbalanced across pairs of subjects.

As our primary concern was to understand the influence of object properties on BOLD activation, we acquired item-wise ratings of our stimuli for 7 different stimulus characteristics. Six of these characteristics (described further below) were putative properties of the objects themselves: Placeness, spatial definition, contextual associations, fixedness, distance, and physical size. The seventh characteristic, visual size was a property of the image (specifically, how many pixels the object took up on the screen). Separate groups of subjects rated these characteristics (except for visual size) either in the laboratory or online. Ratings for object-based characteristics (placeness, spatial definition, contextual associations, fixedness, and physical size) were collected on Object (object + no background) stimuli, while ratings for distance were collected on Scene (object + background) stimuli. Below, we outline the rationale and method for attaining each dimension's ratings, which were ultimately used as regressors in the fMRI analyses. For examples of stimuli with low, medium, and high ratings on each of these dimensions, see Figure 1. Note that these object characteristics were not meant to be orthogonal to each other. Rather, they were chosen based on theoretical considerations and/or inclusion in previous literature, with full understanding that several of them are highly similar.

Placeness

The PPA, RSC, and TOS are usually defined based on greater response to scenes compared with other objects. Although there are many differences between Scenes and Objects that could potentially drive the preferential response to scenes, one of the simplest is that scenes are—conceptually—places, in the sense of being fixed locations in the world. Objects, on the other hand, are not places, but might be considered place-like to varying degrees. For example, a house is an object, because it is a compact single entity, but it is also a place,
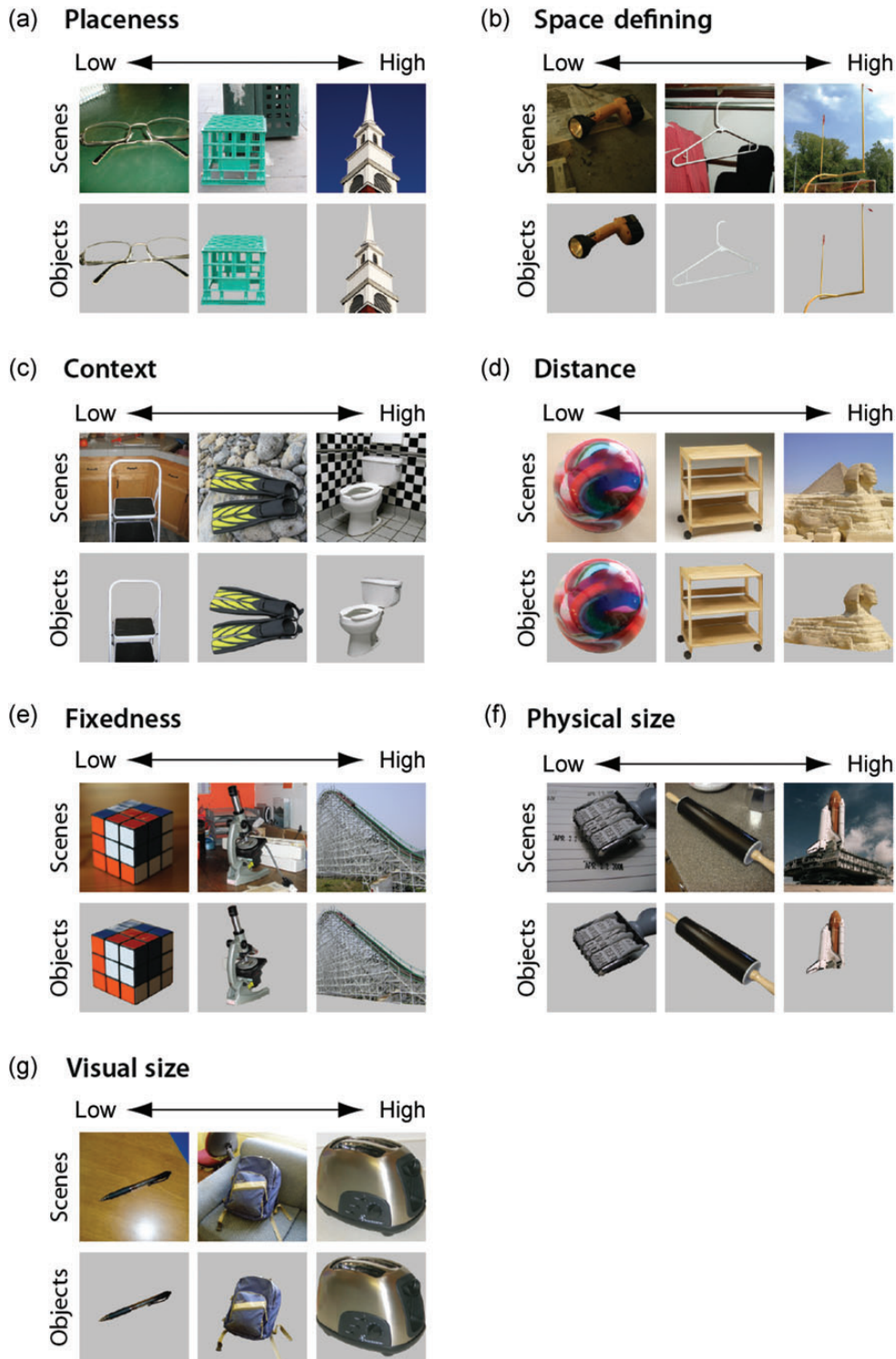
Figure 1. Object properties and example stimuli. Two hundred stimuli were rated by 6 sets of independent raters on 6 stimulus qualities (placeness, spatial definition, contextual associations, distance to object, fixedness, and physical size). Visual size was also calculated based on number of pixels. Ratings were done on object-only stimuli except for the distance to object rating, which was collected on objects with a background.

because it is a stable landmark that might serve as a navigational reference or destination. In a behavioral study ($n = 15$), participants reported whether the contents of each image were more place-like or object-like. The explicit instructions were to report whether each stimulus was more like "a place" or "a thing." A cumulative score for each object was averaged across participants, resulting in scores from 0 (object-like) to 1 (place-like). Note that although we did not define

the terms place and "thing" for the participants, they found this distinction to be intuitive and made the judgments without hesitation.

Spatial Definition
The second object quality assessed was spatial definition; that is, the degree to which an object evokes a sense of surrounding space. This characteristic was originally proposed by Mullally and Maguire (2011)

as a critical driver of the PPA response. The logic here is that some objects (such as a sofa) help define the shape of the space around them, while other objects (such as a bowling ball) provide more ambiguous cues about the shape of surrounding space. The concept of spatial definition is related to the concept of placeness, except that spatial definition focuses explicitly on the physical qualities of the object (for example, whether it is large enough to define the space around it and whether its shape carves out that space), whereas placeness is a broader construct that might conceivably encompass abstract features. To measure this quality in our own stimulus set, we obtained spatial definition ratings in a procedure identical to that outlined in Mullally and Maguire (2011). Briefly, participants ($n = 15$) were instructed to rate objects as space defining (SD) or space ambiguous (SA). SD objects were described as those evoking a strong sense of surrounding space, while SA objects were described as lacking this property. Then for each image, subjects either responded whether the stimuli were 1) SD objects, 2) SA objects, or 3) could not be easily classified. SD responses were given a score of 1, while SA objects were given a score of 0. Objects that could not be easily classified were given a score of 0.5. Each object's cumulative score was averaged across participants to obtain a score between 0 and 1.

Contextual Associations
The third quality examined was the strength of the contextual associations for each object; that is, the degree to which each object was strongly associated with a specific context and to other objects that tend to co-occur in that context. Under this construct, some objects are considered to have strong contextual associations, because they typically appear in the same contexts with certain other objects (for example, a beach ball tends to appear on a beach along with a beach umbrella), while other objects have weak contextual associations, because they may appear in a wide variety of settings with a wide variety of other objects (such as a Rubik's Cube; Bar and Aminoff 2003).

Previous studies of contextual associations have used binary assignments for objects (e.g. "strong context objects" vs. "weak context objects"). However, we reasoned that a binary assignment might not reflect the full range of contextual associations in a large stimulus set. To construct a continuous quantity that corresponded to the specific associations of our stimuli, we calculated a score based on the number and consistency of object-to-context associations for each object. These were determined by first obtaining individual contextual associations through a behavioral experiment, in which participants ($n = 15$) were asked to report the place where they would most typically see each object (following the procedure described in Bar and Aminoff (2003). Three independent raters evaluated the number of unique contexts ascribed to each stimulus (with a broad acceptance for synonymous answers) and the number of subjects who gave each unique context as an answer. This was then converted into a context rating for each stimulus by calculating the entropy, which is defined as $p(x_n) \times \log p(x_n)$ summed over all answers x_n (i.e. types of places, or contexts), with $p(x_n)$ being the probability of answer x_n out of all answers given for that stimulus. For example, if the stimulus in question was a baseball, and we received 3 groups of answers from 15 subjects (9 responded "baseball field," 5 "park," and 1 "playground"), we would convert the number of subjects giving each response into a probability (0.6 for baseball field, 0.33 for park, and 0.07 for playground), and then calculate the entropy as $0.6 \times \log(0.6) + 0.33 \times \log(0.33) + 0.07 \times \log(0.07)$, which is -0.36. For 15 subjects, the range of context values can extend from -1.18 (i.e. the case of weakest contextual association in which each subject named a different setting for the object) to 0 (i.e. the case of strongest contextual association in which all 15 subjects named the same setting). Note that an advantage of using the entropy measure, as opposed to simply counting the number of distinct answers given by the set of participants, is that it includes information about the consistency of these answers. For example, the entropy score will be higher for an object for which 14 of the 15 subjects gave the same answer and 1 of the 15 gives a different answer than for an object for which 7 of the 15 subjects gave 1 answer and 8 of the 15 gave another, even though the total number of distinct answers is the same in both cases.

Fixedness
An important characteristic of objects that may contribute to their role in scene processing is how fixed they are in the environment. For instance, a completely immobile landmark, such as a monument, may be represented as a scenic background element and activate PHC to a greater extent than an object that is more easily transportable. Furthermore, fixed objects are more useful as navigational referents than objects that change their location. Notably, a recent study reported that PHC and retrosplenial cortex responded more strongly to fixed than to nonfixed objects (Auger et al. 2012). To assess fixedness, subjects ($n = 20$) were asked to report how easily they could pick up and move each object on a scale of 1–5 (with a rating of 1 corresponding to an easily movable object and a rating of 5 corresponding to an immobile or difficult to move object). Ratings were then averaged for each stimulus, with values ranging from 1 to 5 (mean = 2.16).

Physical Size
Both physical and perceived object size have been found to influence activity levels in the PHC. To assess the contribution of this object quality, subjects ($n = 21$) were instructed to estimate the objects' length (in feet) along their longest dimensions. Responses were averaged for each object, resulting in values ranging from 0.063 to 332.69 feet (mean = 8.45). These values were then log-scaled.

Distance
Perception of the distance of an object to oneself is critical for interacting with our environment. Previous work has demonstrated that more distal objects activate the PPA to a greater extent than proximal objects (Amit et al. 2012). To determine the perceived distance to our object stimuli, we collected ratings in a behavioral experiment ($n = 24$). Because it is difficult to assess the distance of an object without reference to other elements of the scene, these ratings were made on the Scene (object + background) stimuli. Thus, subjects were asked to rate the distance of the object as shown in that particular image; however, we expect that these ratings would correspond to the typical distance at which an object is viewed, which is a property of the object itself. Participants rating distance were instructed to estimate the distance (in feet) to each focal object. Responses ranged from 0.414 to 341.30 feet (mean = 9.05). Responses for each object were then averaged and log-scaled.

Visual Size
Finally, the number of pixels in each focal object was calculated with Adobe Photoshop. Because the screen in the scanner was fixed at the same distance for all subjects, this measure corresponds to the retinotopic size of the objects. Values ranged from 9851 to 433 237 pixels (mean = 170 101).

Inter-rater reliability metrics were calculated in 1 of 2 ways. For object qualities resulting in a binary vector for each participant (placeness and spatial definition), we calculated the consistency of individuals across each item as a percentage. For example, if 14 of the 15 individuals had the same response on a particular item, the reliability for that item is 0.93. These values were then averaged across all items. Reliability was 0.82 for placeness and 0.64 for spatial definition. Because contextual associations calculated using an entropy score are themselves a measure of rater consistency (with a low context object having low consistency and a high context object having high consistency), an inter-rater consistency score is not appropriate for this metric. For all other object qualities, inter-rater reliability was calculated by computing the average correlation between each pair of raters and then averaging these correlation values. Mean inter-rater consistency for fixedness, physical size, and distance were 0.81, 0.62, and 0.64, respectively.

Procedure
The main fMRI experiment consisted of two 7 min 30 s scan runs, each of which was divided into 50 trials in which subjects saw Scenes (object + background), 50 trials in which they saw Objects (object + no

background), and 50 null events. During each experimental trial, subjects viewed a Scene or Object for 1.5 s followed by a 1.5-s blank period. They were instructed to press a button for each stimulus occurrence; beyond this, their instructions were simply to attend to each stimulus, including both foreground object and background elements (when present), and memorize it for a subsequent memory test. During 3-s null events, no stimulus appeared; these served to jitter the interval between experimental trials. A fixation cross remained on the screen throughout the entire experiment and subjects were asked to keep their eyes on the cross even when attending to information on other parts of the screen. Trial ordering was optimized using Optseq2 (http://surfer.nmr.mgh.harvard.edu/fswiki/optseq2) to maximize power for detecting differences between Scenes and Objects; a different trial ordering was used for each subject. Every subject saw every focal object once in the course of the experiment, either with or without a background, with a paired subject seeing the complementary version.

Following the main experimental scans, 2 functional localizer scans were administered, in which subjects made 1-back repetition detection judgments on scenes, faces, objects, and other stimuli, presented in a blocked design. These stimuli had not been previously seen in the main experiment.

Outside of the scanner, subjects were subsequently given a memory test in which they viewed each focal object, presented without background, along with 50–75 previously unseen lures. For each item, they were asked to report whether 1) they had seen it in the scanner with a background, 2) they had seen it in the scanner without a background, and 3) they had not seen it before. Subjects correctly reported previously seen items as being familiar (hits) on 72% of trials and incorrectly reported previously unseen lures as being familiar (false alarms) on 17% of trials ($d' = 1.84$, $t_{11} = 7.1$, $P = 0.000019$). For items judged to be familiar, subjects correctly reported background presence/absence on 68% of trials (chance = 50%, $t_{11} = 3.6$, $P = 0.004$). Thus, subjects encoded both the identities of the objects and also information about whether the objects were presented with background or not, suggesting that they followed the instructions to pay attention to the stimuli during the scan. Note that this was despite the fact that subjects were not aware beforehand of the precise nature of this memory test. In particular, they had no foreknowledge that stimuli would be presented without backgrounds during this phase and, thus, had no reason to focus on the objects and/or backgrounds in particular rather than the scenes as a whole.

Data Analysis

Functional images from both experimental and localizer scan runs were corrected for differences in slice timing by re-sampling slices in time to match the first slice of each volume, spatially realigned with respect to the first image of the scan to compensate for head motion, spatially normalized to the Montreal Neurological Institute template, and spatially smoothed with a 6-mm full-width at half-maximum Gaussian filter. Data were analyzed in VoxBo using a general linear model (GLM) including filters that removed high and low temporal frequencies, an empirically derived $1/f$ noise model, regressors to account for global signal variations, and nuisance regressors to account for between-scan differences.

Five functional regions of interest (fROIs) were defined in each subject using data from the functional localizer scans. The PPA (average size = 104 voxels in each hemisphere, range 66–209 voxels) was defined as the set of contiguous voxels responding more strongly to scenes than to objects in the posterior parahippocampal/collateral sulcus region. To ensure that our findings were not a result of restricting analyses to only the most scene-selective voxels in the PHC, we also implemented an anatomically and functionally constrained definition of the PHC in which all voxels within the PHC as defined by Pruessner et al. (2002) for which activation was higher than baseline ($t > 2.0$) for either Scenes or Objects during the main experimental scans were included, and so we only report findings from the union of voxels in each subject's functionally/anatomically defined PHC and functionally defined PPA (referred to as PPA/PHC, average size = 146 voxels in

each hemisphere, range 79–278 voxels). The RSC (average size = 75 voxels in each hemisphere, range 37–237) was defined as the set of contiguous voxels in the retrosplenial/medial parietal region responding more strongly to scenes than to objects. The TOS (average size = 93 voxels in each hemisphere, range 56–230) was defined as the set of contiguous voxels responding more strongly to scenes than to objects near the sulcus of the same name. The lateral occipital complex (LOC, average size = 340 voxels in each hemisphere, range 101–826) was defined as the set of contiguous voxels responding more strongly to objects than to scrambled objects in the lateral portions of occipitotemporal cortex. Finally, early visual cortex (EVC, average size = 192 voxels in each hemisphere, range 45–380) was defined as the set of contiguous voxels that responded more strongly to scrambled objects than to objects in the posterior visual cortex. Significance thresholds (ranging from $t > 3.0$ to $t > 4.0$) were set for each ROI on a subject-by-subject basis to be consistent with those identified in previous studies (Epstein and Kanwisher 1998; Epstein et al. 1999; Epstein and Higgins 2007). The time course of response during the main experiment was then extracted from each ROI and subjected to several analyses, described below.

The first set of analyses examined the modulation of fMRI activity by stimulus characteristics using a standard linear modeling approach. Because the 7 stimulus characteristics tended to be highly correlated with each other (Table 1), making it difficult to assign unique variance to each, we investigated each one using a separate model. These models included 2 categorical regressors indicating whether the stimulus on each trial was a Scene (object + background) or an Object (object + no background), along with 2 continuous-value covariates indicating the trial-wise values of the targeted stimulus characteristic, for Scene and Object trials separately. To facilitate the comparison of the magnitude of effects across characteristics, covariate values were z-scored (separately for Scenes and Objects) for each subject before analysis. Note that the covariate values were identical for the corresponding Scene (object + background) and Object (object + no background) versions of each stimulus prior to z-scoring, irrespective of whether the ratings were originally obtained on with-background or without-background stimuli. Categorical regressors were impulse response functions at stimulus onset convolved with the VoxBo canonical hemodynamic response function (HRF; which is the first eigenvector of a population of HRFs obtained from motor cortex in a previous study see Aguirre et al. 1998). Continuous-value covariates were impulse response functions scaled appropriately for each trial and then convolved with the same canonical HRF. Response estimates (i.e. β-values) were obtained for each regressor and covariate, which were then compared with zero using a 2-tailed t-test. Whole-brain analyses were also implemented for these models by applying the same GLM to every voxel of the brain and by entering the resulting subject-specific β-maps into a second-level random effects comparison against zero (i.e. no effect). We used a relaxed threshold ($P < 0.05$, uncorrected) to visualize the results of these analyses on whole-brain maps (Fig. 3). It should be noted that voxels identified at this threshold could be Type I errors.

Using a similar approach, we also examined BOLD response to variance components that were shared across several of the 7 stimulus

dimensions. To determine these components, we first submitted the item-wise behavioral ratings for the 7 parameters to a principal components (PC) analysis. A varimax rotation was used to derive orthogonal factors with an eigenvalue >0.75. After completing the factor analysis to identify shared underlying components, we estimated the coefficients for each of the factors for each stimulus and entered these as regressors of interest in a GLM fMRI analysis. Region of interest and whole-brain analyses were then implemented as previously described.

The second set of analyses (item analyses) also examined the modulation of fMRI activity by stimulus characteristics, but in a different manner. Rather than directly modeling the parametric modulation of effects on the BOLD time course, we used a 2-stage procedure, in which we first estimated the BOLD response for each item and then examined the effect of stimulus characteristics on these item-wise responses in a second-level analysis. The GLM used in the first stage to estimate these item-wise responses included a single regressor for each trial of the experiment (200 in all), modeled as an impulse response function convolved with a standard hemodynamic response function. Individual β-values were then z-scored for each ROI for each subject before averaging across subjects to obtain a response estimate for each item. (The z-score approach was necessary to normalize item effects across subjects, because not every stimulus was seen by every subject; specifically, each subject saw either the with-background or without-background version of each object, but not both.) A series of linear and step-wise regression analyses were then implemented in SPSS 19 to determine the effect of the 7 stimulus characteristics on the item-wise response values (described further in the Results, below). These analyses were performed separately for the Scene and Object stimuli.

There are 2 advantages to this 2-stage, item-wise approach. The first is that the BOLD response to each trial is estimated without reference to a cognitive hypothesis; thus, the accuracy of these estimates is not affected by colinearities among the cognitive quantities of interest (in this case, the various stimulus characteristics). This is especially advantageous because colinearities between regressors are typically accentuated by convolution with the hemodynamic response function. The second advantage is that the reliability of effects is determined by reference to the variability across items—rather than across subject, as is more typical—thus allowing generalization of the conclusions to other items not shown.

Results

Response to Scenes versus Objects

The PPA, RSC, and TOS are customarily defined based on greater response to scenes than to objects. An important difference between these stimulus classes is that scenes have background elements, whereas objects do not. Half of the stimuli in the current experiment were objects with background (i.e. Scenes), while half were objects without background (i.e. Objects). Before examining activity relating to the characteristics of the focal object, we first examined activity relating to the presence or absence of a scenic background.

As expected the 3 scene-selective ROIs (PPA/PHC, RSC, and TOS) responded much more strongly to with-background stimuli (i.e. Scenes) than to without-background stimuli (i.e. Objects; PPA/PHC: $t_{11} = 21.36$, $P = 0.00000000003$; RSC: $t_{11} = 6.77$, $P = 0.00003$; TOS: $t_{11} = 7.03$, $P = 0.00002$). Given that the focal objects were identical for both stimulus sets, this suggests that the mere presence or absence of a scenic background can be a strong determinant of response levels in these regions. The EVC also responded more strongly to Scenes than to Objects ($t_{11} = 6.27$, $P < 0.001$). This result is expected, because the Scenes cover a larger portion of the visual field than the Objects and, thus, should activate a larger portion of retinotopic cortex. The LOC also showed a trend for greater response to Scenes, but this trend fell short of significance (LOC; $t_{11} = 2.13$, $P = 0.056$).

Response to Individual Object-Based Properties

We then turned to the main question of the study: To what extent is activity in scene regions modulated by object-based properties? To address this, we examined the response to each property in turn, considering in each case the response to objects with background (i.e. Scenes) and objects without background (i.e. Objects) separately (Table 2). Note that this approach did not allow us to distinguish effects that were unique to a property from effects that were shared with other properties—we address the issue of distinct versus shared components in the subsequent analyses.

Consistent with the results of previous studies, PPA/PHC was sensitive to all 6 object-based properties (Fig. 2), showing significant or near-significant modulation to each when background was present (all $Ps < 0.05$, except fixedness $P = 0.07$) and significant modulation to all properties except context when background was absent (all $Ps < 0.05$, except context $P = 0.13$). Response in PPA/PHC was also affected by the visual size of the object, but also when objects were shown in isolation ($P = 0.008$), not only when they were shown with a scenic background ($P = 0.34$, ns). This last set of results is consistent with previous reports that PPA shows retinotopic organization (Arcaro et al. 2009; Ward et al. 2010), as isolated objects that cover more screen pixels will activate a larger retinotopic extent, whereas the object + background stimuli all activate the same retinotopic extent irrespective of the visual size of the focal object.

Similar results were observed in RSC, which was also sensitive to all 6 object-related properties when background was present (all $Ps < 0.01$) and to all object-related properties except context when background was absent (all $Ps < 0.01$, except context $P = 0.22$). There was no effect of visual size for either Scenes or Objects (both $Ps > 0.12$, ns) consistent with earlier reports (e.g. Ward et al. 2010) that RSC shows little evidence of retinotopic organization. TOS was sensitive to 4 out of 6 object properties (placeness, spatial definition, fixedness, and physical size) when the objects were shown in isolation (all $Ps < 0.05$), but was not sensitive to any object property when objects were shown as part of a scene ($Ps > 0.08$, ns). As with the PPA/PHC, an effect of visual size was observed for Objects ($P = 0.012$), but not Scenes ($P = 0.48$, ns).

Next, we compared the 6 object-based properties to each other (setting visual size aside for the moment). These properties are similar; however, if one of them is a better approximation of the "true" factor (or factors) that drives a region, then it should modulate fMRI response more strongly. To test this, we performed a 6×2 analysis of variance (ANOVA) for each region, with stimulus property and background presence (Scene vs. Object) as factors, to determine (1) whether the 6 object-based properties modulated fMRI response to different degrees, (2) whether the strengths of the object-based property effects varied depending on whether background was present or not.

In the PPA/PHC, the main effect of object-based property was highly significant ($F_{5,55} = 6.6$, $P = 0.00007$), indicating that PPA/PHC response is better predicted by some object-based

Table 2
Results of GLMs in fROIs

	Scenes (object + background)				Objects (object + no background)			
	% Change	SEM	t	P-values	% Change	SEM	t	P-values
PPA/PHC								
Placeness	0.067	0.016	4.247	0.001**	0.043	0.012	3.707	0.003**
Space defining	0.053	0.012	4.466	0.001***	0.043	0.012	3.547	0.005**
Context	0.036	0.012	3.095	0.010*	0.013	0.008	1.619	0.134
Distance	0.039	0.014	2.728	0.020*	0.031	0.010	2.997	0.012*
Fixedness	0.031	0.015	2.038	0.066	0.045	0.011	4.251	0.001**
Physical size	0.039	0.014	2.870	0.015*	0.043	0.012	3.447	0.005**
Visual size	0.010	0.010	1.003	0.337	0.043	0.013	3.241	0.008**
RSC								
Placeness	0.078	0.018	4.423	0.001**	0.045	0.010	4.578	0.001***
Space defining	0.076	0.017	4.358	0.001**	0.041	0.011	3.717	0.003**
Context	0.049	0.013	3.869	0.003**	0.009	0.007	1.299	0.220
Distance	0.086	0.014	5.995	0.000***	0.050	0.013	3.699	0.004**
Fixedness	0.060	0.017	3.486	0.005**	0.046	0.012	3.736	0.003**
Physical size	0.081	0.014	5.596	0.000***	0.054	0.011	5.040	0.000***
Visual size	0.015	0.010	1.598	0.138	0.001	0.015	0.082	0.936
TOS								
Placeness	0.045	0.018	2.543	0.027*	0.054	0.018	3.066	0.011*
Space defining	0.030	0.016	1.899	0.084	0.074	0.013	5.532	0.000***
Context	0.009	0.015	0.628	0.543	0.026	0.016	1.620	0.133
Distance	0.018	0.021	0.871	0.403	0.030	0.017	1.795	0.100
Fixedness	0.006	0.016	0.384	0.709	0.057	0.014	4.060	0.002**
Physical size	0.012	0.018	0.645	0.532	0.052	0.017	3.082	0.010*
Visual size	0.014	0.019	0.726	0.483	0.059	0.020	2.979	0.013*
LOC								
Placeness	−0.030	0.016	1.842	0.093	0.019	0.016	1.170	0.267
Space defining	−0.035	0.016	2.186	0.051	0.014	0.015	0.896	0.390
Context	−0.031	0.018	1.701	0.117	0.004	0.016	0.256	0.803
Distance	−0.034	0.021	1.645	0.128	−0.012	0.016	0.754	0.467
Fixedness	−0.046	0.017	2.664	0.022*	−0.001	0.013	0.079	0.938
Physical size	−0.043	0.017	2.490	0.030*	−0.024	0.019	1.237	0.242
Visual size	0.006	0.013	0.440	0.669	−0.015	0.015	0.967	0.354
EVC								
Placeness	−0.009	0.014	0.610	0.554	−0.016	0.029	0.545	0.597
Space defining	−0.007	0.016	0.468	0.649	0.005	0.028	0.165	0.872
Context	−0.006	0.010	0.664	0.521	−0.007	0.032	0.205	0.841
Distance	−0.003	0.021	0.137	0.894	−0.007	0.029	0.250	0.807
Fixedness	−0.011	0.025	0.436	0.672	0.027	0.027	1.000	0.339
Physical size	−0.004	0.023	0.194	0.850	0.010	0.038	0.273	0.790
Visual size	0.010	0.020	0.500	0.627	0.123	0.036	3.422	0.006**

*$P < 0.05$.
**$P < 0.01$.
***$P < 0.001$.

characteristics than by others. Inspection of the results (Fig. 2) indicates that placeness and spatial definition had the strongest effect on PPA response, whereas context had the weakest effect. There was no main effect of background presence ($F < 1$, ns), indicating that object-based properties modulated the PPA equally well for Scene and Object stimuli. There was also an interaction between property and background presence ($F_{5,55} = 2.8$, $P = 0.024$), most likely reflecting the fact that the context effect was stronger for Scenes than for Objects, while the fixedness effect was stronger for Objects than for Scenes.

RSC also showed differential modulation by object property ($F_{5,55} = 15.9$, $P = 0.000000001$). Here, the strongest effect was due to placeness, with distance and physical size not far behind, followed by spatial definition. Again, context had the weakest effect. RSC showed greater modulation by object properties for Scenes than for Objects ($F_{1,11} = 6.09$, $P = 0.031$), and there was an interaction between property and background presence ($F_{1,11} = 7.17$, $P = 0.00003$), most likely reflecting the stronger effect of context for Scenes than for Objects.

TOS also showed differential modulation by object property ($F_{5,55} = 5.1$, $P = 0.001$). As with the PPA, placeness and spatial definition had the strongest effect, whereas context had the weakest effect. Object-property effects did not differ in size for Scene and Object stimuli ($F_{1,11} = 2.7$, $P = 0.13$), a surprising result in light of the previous analyses which suggested that these effects were only significant for Objects. Inspection of the data revealed that although this pattern was indeed observed in 11 of the 12 subjects, there was a single subject who exhibited the opposite pattern. When this subject was removed, TOS also showed an effect of background presence ($F_{1,10} = 23.4$, $P < 0.001$), indicating that object-property effects were more salient when objects were shown in isolation. There was no interaction between object property and background presence.

In sum, scene-responsive regions are sensitive to object-based properties; furthermore, some object-based properties modulate responses in these regions more strongly than others. To see if the object-property effects are general across visual cortex, we examined response in 2 control regions: The LOC and EVC. There was a trend for LOC to respond
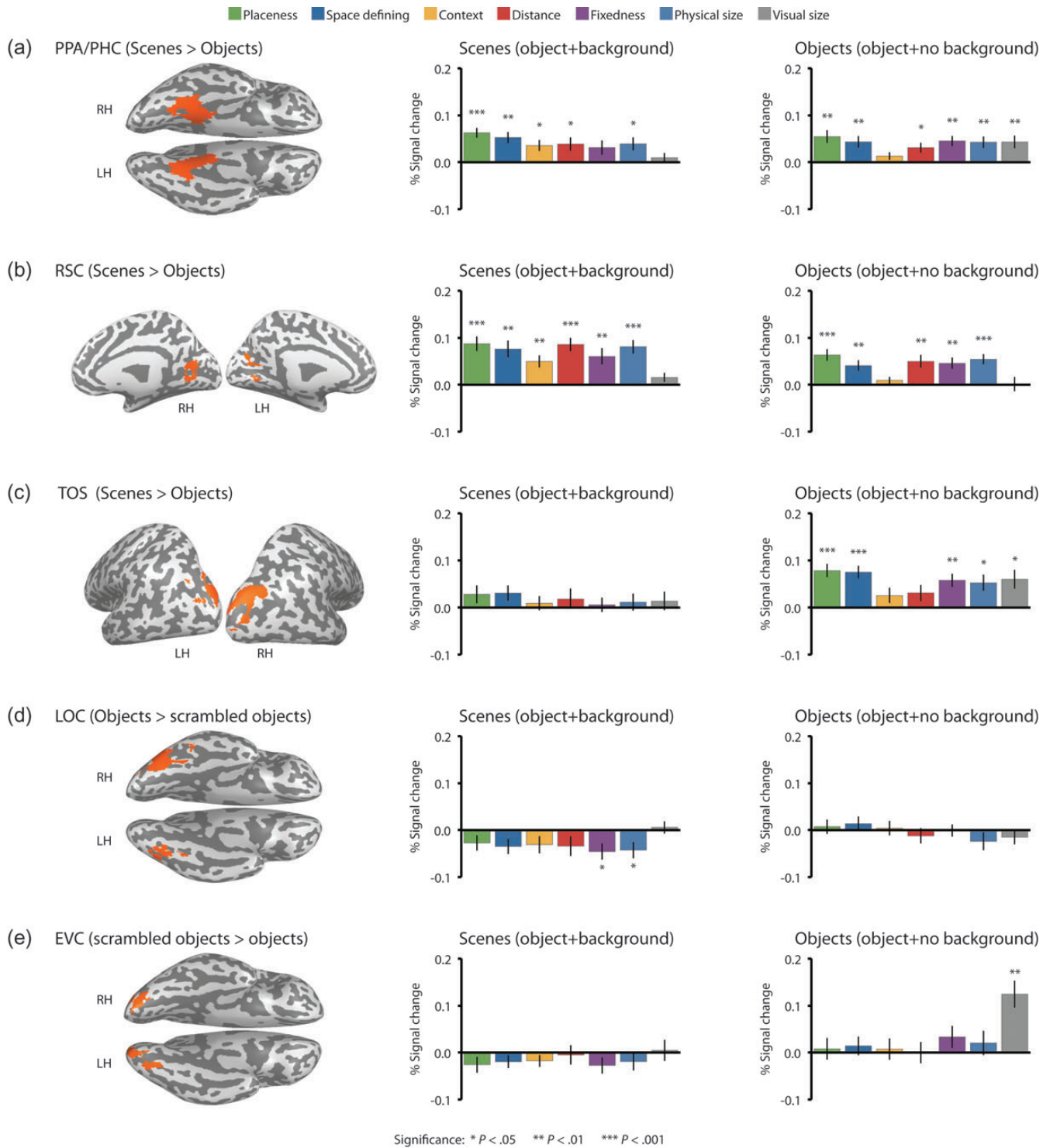
Figure 2. Effects of object properties in scene-selective cortex. fMRI response for each stimulus property within bilateral (*a*) PPA/PHC, (*b*) RSC, (*c*) TOS, (*d*) LOC, and (*e*) EVC for Scenes (object + background) and Objects (object + no background). The PPA/PHC, RSC, and TOS were defined based on a scene > object contrast; LOC was defined based on an intact object > scrambled object contrast; EVC was defined based on a scrambled object > intact object contrast. The majority of stimulus properties modulated response in the PPA/PHC and RSC for both Scene and Object stimuli. In contrast, stimulus properties modulated the TOS only when objects were presented without a background. LOC showed a weak inverse sensitivity to 2 of the stimulus properties. EVC was sensitive to the number of pixels of the focal object, but not to other factors.

negatively to object-based properties (i.e., to respond more weakly to objects that score high on these properties) when background was present (Fig. 2). This trend was significant for fixedness and physical size ($P < 0.05$) and marginal for spatial definition ($P = 0.051$). No modulation by object-based properties was seen when objects were shown in isolation (all $Ps > 0.2$, ns) and (somewhat surprisingly) no effects of visual size were observed.
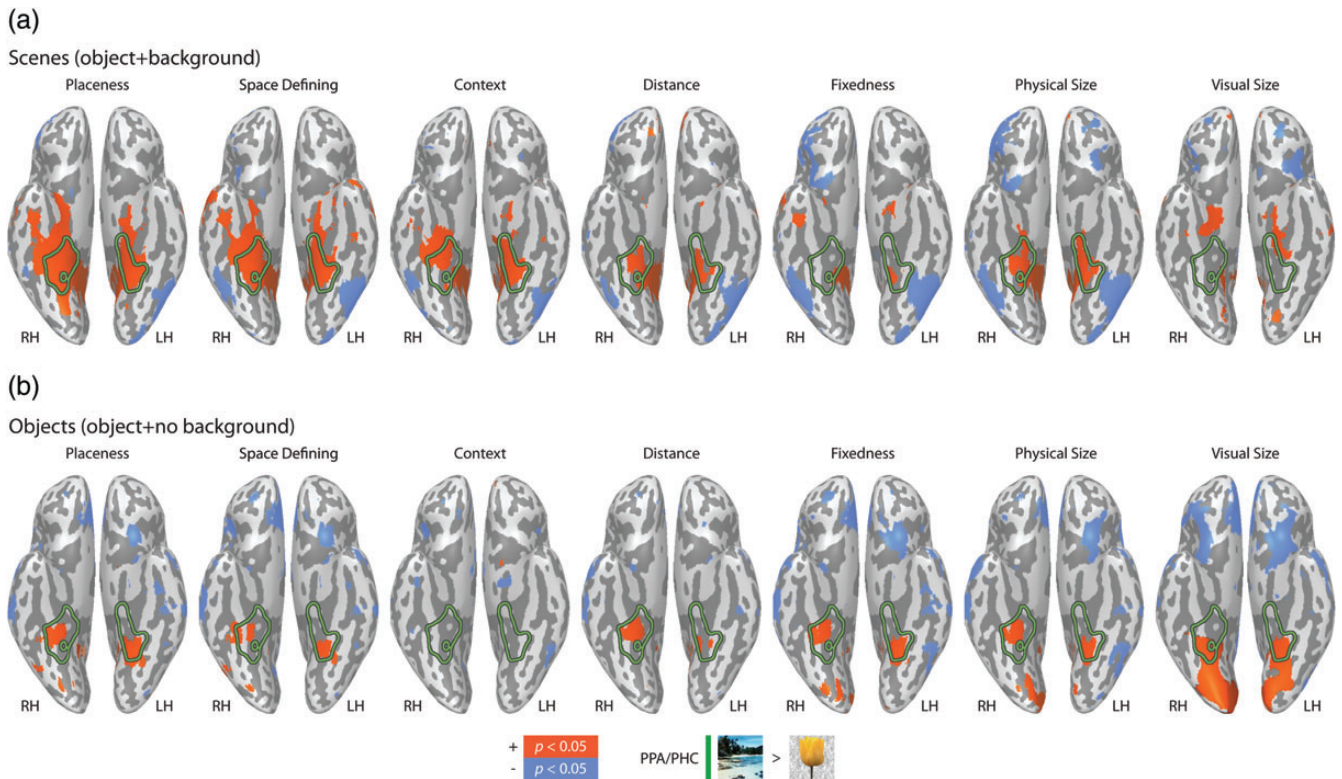
Figure 3. Data from whole-brain analysis showing effects of object properties in ventral occipitotemporal cortex. Voxels showing increasing activation with greater values of each property are in orange, while voxels showing the reverse pattern (decreasing activation with greater values of each property) are in blue, for (*a*) Scenes (objects + background) and (*b*) Objects (object + no background). PPA/PHC from the localizer data is outlined in green. This region is commonly activated for most of the 6 object properties. Sensitivity to visual size (i.e. retinotopic extent) was observed for Objects in PPA/PHC and also in more posterior visual regions. Note that a $P < 0.05$ uncorrected threshold is used in this figure to emphasize that object-based effects in ventral occipital cortex are consistent within the PPA/PHC but unreliable outside; because of the liberal threshold, these maps are likely to include Type I errors.

As expected, EVC was insensitive to object-based properties (all $Ps > 0.3$, *ns*) and was highly sensitive to the visual size of objects when they were shown in isolation ($P = 0.001$), but not when they were shown as part of a scene ($P = 0.33$, ns). Because responses to object-based properties were weak in LOC and EVC, we did not perform further analyses on these responses.

Results of a whole-brain analysis generally confirmed the findings of the ROI analysis (Fig. 3). In particular, the PPA/PHC seems to be a central locus that is activated by the 6 object-based properties and also by visual size. We also observed a trend toward the opposite sensitivity in the region lateral to the PPA/PHC corresponding to LOC, but only for Scene stimuli.

Response to Components Shared Across Object-Based Properties

We next turned to an examination of BOLD response to hypothetical stimulus dimensions that might be shared across several of our original object-based properties. The idea that there might be a smaller number of fundamental dimensions is a compelling one, because the characteristics in question tend to covary with each other. For example, a sofa is a large object, which tends to be fixed in place, associated with a certain kind of place (a living room), and so on. Indeed, the various factors were all related to some degree in our stimulus set (Table 1). Specifically, placeness, spatial definition,

Table 3.
Principal Components Analysis.

	Factor 1 ("Landmark Suitability")	Factor 2 ("Visual Size")	
Placeness	0.866*	0.101	
Space Defining	0.898*	0.017	
Context	0.631*	−0.303	
Distance	0.875*	−0.052	
Fixedness	0.912*	−0.175	
Physical Size	0.885*	0.217	
Visual Size	0.084	0.955*	
			Total Variance
Variance(%)	*62.4*	*15.26*	*77.69*

* indicates highest factor loading

context, distance, fixedness, and physical size were all highly correlated with each other (although context was somewhat less well-correlated than the other factors). Visual size, on the other hand, was only weakly correlated with physical size and fixedness, and not correlated at all with the other properties.

The structure in the stimulus space was confirmed by a PC factor analysis, which revealed 2 underlying factors (Table 3). The first factor corresponded to the physical, semantic, and associative aspects of the objects as revealed by the 6 object-based characteristics. The second factor corresponded closely to visual size. These results suggest that the variability in our stimulus set might be largely explained by a single object-
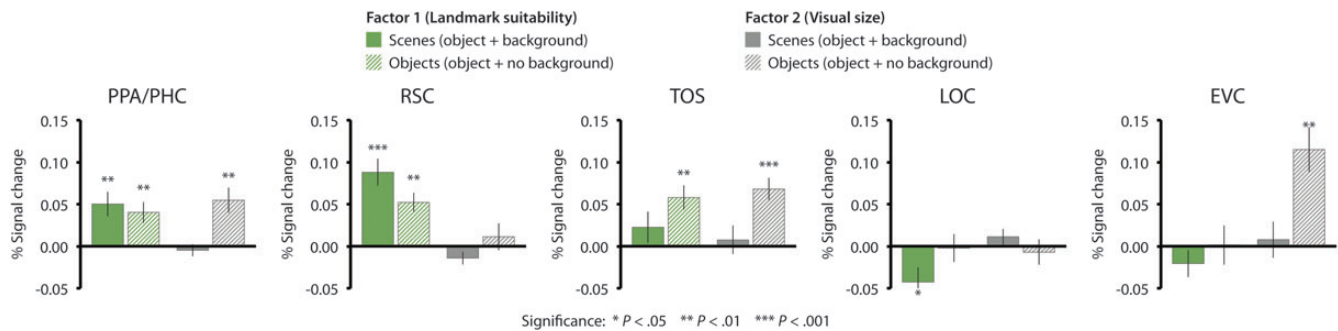
Figure 4. fMRI response to the 2 PC of the item-wise ratings of stimulus properties in PPA/PHC, RSC, TOS, LOC, and EVC. fMRI response to factor 1 (landmark suitability) is plotted in green, while response to factor 2 (visual size) is plotted in gray. Solid bars indicate response to Scene (object + background) stimuli, while striped bars indicate response to Object (object + no background) stimuli. The PPA/PHC and RSC were sensitive to factor 1 for both Scenes and Objects, while TOS was sensitive to this factor for Objects only. Retinotopic responses (i.e. sensitivity to factor 2 for Objects only) were observed in PPA, TOS, and EVC, but not RSC or LOC.

based feature dimension (in addition to visual size). We label this factor the "landmark suitability" of the object, because the 6 object-based characteristics that compose the factor all relate in some way to the appropriateness of the object as a navigational landmark.

To confirm that the 2 factors (landmark suitability and visual size) engaged scene regions in a manner similar to individual object characteristics, we ran a GLM with the 2 independent factors scores from the factor analysis (Fig. 4). PPA/PHC was sensitive to the first factor (landmark suitability) for both Scenes and Objects (Scene: $t_{11} = 3.4$, $P = 0.005$; Object: $t_{11} = 3.3$, $P = 0.007$), but was only sensitive to the second factor (visual size) for Objects (Scene: $t < 1$, ns; Object: $t_{11} = 3.7$, $P = 0.004$). RSC was sensitive to the landmark suitability factor for both Scenes and Objects (Scene: $t_{11} = 5.5$, $P = 0.0001$; Object: $t_{11} = 4.7$, $P = 0.001$), but was insensitive to the visual size factor in both cases (Scene: $t_{11} = -1.8$, $P = 0.09$, ns; Object: $t < 1$, $P = 0.70$, ns). TOS was sensitive to both landmark suitability and visual size, but for Objects only (factor 1: Scene: $t_{11} = 1.2$, $P = 0.25$, ns; Object: $t_{11} = 4.0$, $P = 0.002$; factor 2: Scene: $t < 1$, $P = 0.67$, ns; Object: $t_{11} = 5.18$, $P < 0.0003$). Thus, the results of these analyses confirmed the previous observation that both PPA and RSC are sensitive to object-based properties (factor 1) for both Scenes and Objects, whereas TOS is only sensitive to these properties for Objects. Retinotopic effects (factor 2 for Objects), on the other hand, were observed in PPA and TOS, but not RSC.

To further understand how the factors from the PC analysis (landmark suitability and visual size) were modulated by background presence, we performed a 2 × 2 ANOVA in each region, with background presence (Scene vs. Object) and PC factors (landmark suitability and visual size) as factors. In the PPA/PHC, there was no main effect of background presence or PC factor, but a significant interaction effect ($F_{1,11} = 19.8$, $P = 0.001$), reflecting the fact that object-based properties (factor 1) modulated activity for both Scenes and Objects, but visual size (factor 2) only modulated activity for Objects. RSC showed no main effect of background presence, but did show a main effect of PC factor ($F_{1,11} = 22.2$, $P = 0.001$) and a significant interaction between PC factor and background presence ($F_{1,11} = 10.1$, $P = 0.009$), reflecting the fact that this region was sensitive to object properties (factor 1), but not visual size (factor 2) and the fact that these object-property effects were larger for Scenes than for Objects. In the TOS, there was a main effect of background ($F_{1,11} = 6.03$, $P = 0.032$), but no

main effect of PC factor or an interaction effect, confirming the previous observation that this region is sensitive to both factors for Objects, but insensitive to both factors for Scenes.

For completeness, we also consider response to the 2 factors in LOC and EVC. LOC was inversely sensitive to factor 1 for Scenes ($t_{11} = -2.41$, $P = 0.03$), but not Objects ($t_{11} = 0.14$, $P = 0.89$, ns), and was insensitive to factor 2 in both cases (Scene: $t_{11} = 1.16$, $P = 0.270$, ns; Object: $t_{11} = 0.471$, $P = 0.65$, ns). EVC was not sensitive to factor 1 for either Scene or Objects (Scene: $t_{11} = 0.60$, $P = 0.56$, ns; Object: $t_{11} = 0.29$, $P = 0.78$, ns), but was sensitive to factor 2 for Objects, but not for Scenes (Scene: $t_{11} = 0.45$, $P = 0.66$, ns; Object: $t_{11} = 3.77$, $P = 0.001$). Results of a whole-brain analysis generally confirmed the ROI analysis, with PPA and RSC responding to factor 1, TOS responding to both factors for objects only, and EVC responding to factor 2 for objects only (Supplementary Fig. S1).

Item Analysis

The analysis above suggests that the 2 factors revealed by the PC analysis of the item-wise property ratings—corresponding to landmark suitability and visual size—explain a substantial amount of the fMRI response. However, this leaves open the question of whether 2 factors are sufficient to explain the response, or whether individual properties might make independent contributions. To address this question, we turned to an item-wise analysis of the fMRI data. For this analysis, the response to each object was independently assessed (separately for Scene and Object stimuli) in the functional ROIs defined previously. These item-wise responses were then averaged across subjects and compared with the item-wise behavioral ratings.

We first examined the extent to which each of the stimulus properties explained item-wise differences in the fMRI responses (Fig. 5). This was tested by calculating the correlation between item-wise fMRI responses and item-wise ratings for each property. This analysis is analogous to the testing of individual properties described above. However, whereas the previous analysis examines reliability across subjects, here, we examine reliability across stimuli. The conclusion one draws from a significant result in the former case (i.e. standard analysis) is that the finding would replicate in a different group of subjects; the conclusion one draws in the
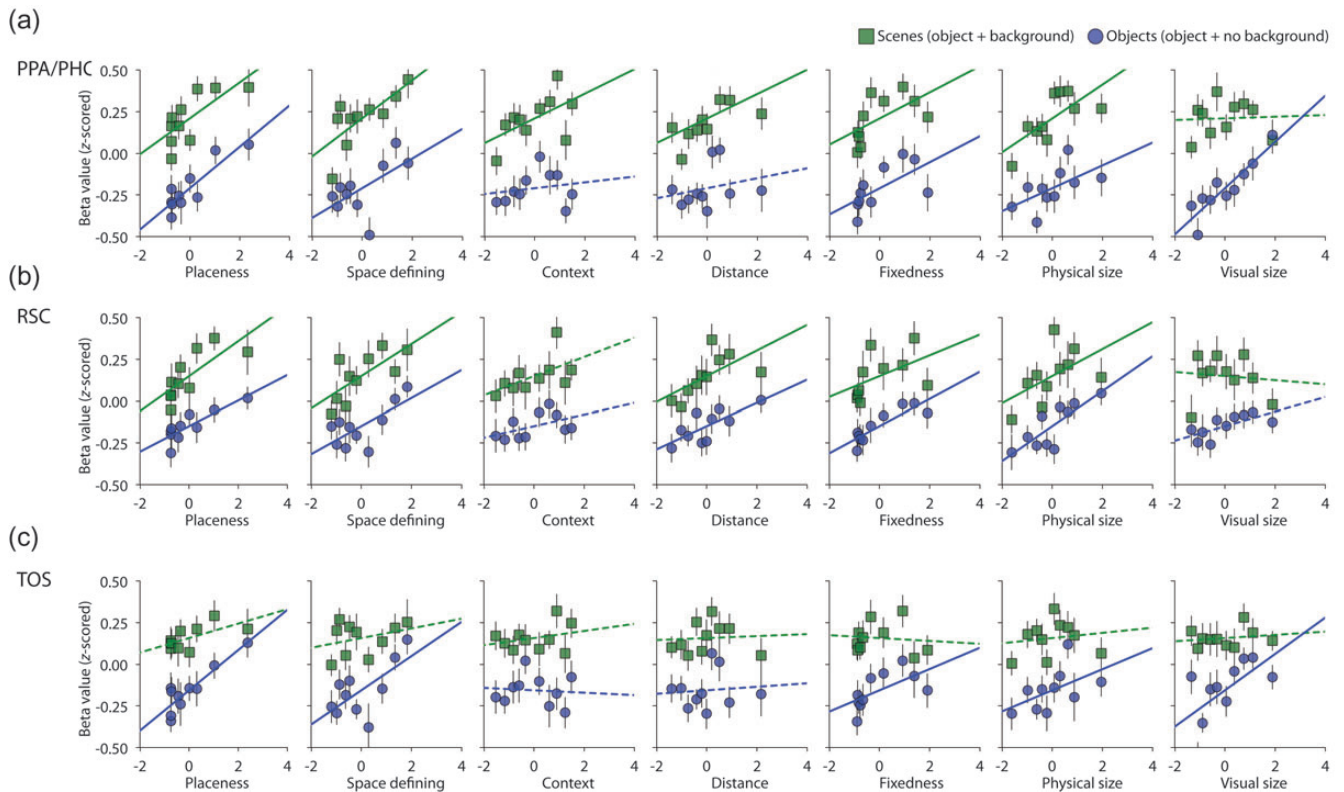
Figure 5. Results of item analysis in scene-selective cortex showing fMRI response plotted as a function of object characteristics in (a) PPA/PHC, (b) RSC, (c) TOS. For purposes of display, items were grouped into sets of 20 based on their property scores and average fMRI response for each group was plotted against the average property score. Responses to Scenes (object + background) indicated by squares and response to Objects (object + no background) indicated by circles. Positive slope reflects the modulation of the region by that particular stimulus property. Solid trend lines indicate a significant effect. Consistent with previous results, we find that PPA/PHC and RSC are modulated by object properties for both Scenes and Objects, whereas the TOS encodes only some of these properties and only for Objects (i.e. isolated objects without scenic background).

latter case (i.e. item analysis) is that the finding would replicate in a different group of items.

The results of the item analyses were consistent with the results of the standard analyses (Table 4). Significant correlation was observed between fMRI activity level in the PPA/PHC and item-wise ratings for 6 of 6 object properties when background was present (all $Ps < 0.05$) and 4 of 6 object properties when background was absent ($Ps < 0.05$, except context and distance, $P > 0.3$, ns). Similar results were observed in RSC: All object properties except context predicted item-wise activity levels, both for Scenes (all $Ps < 0.05$, except context $P = 0.11$, ns) and for Objects (all $Ps < 0.05$, except context $P = 0.26$, ns). Effects for TOS were also similar to the standard analysis: TOS was sensitive to 4 of the 6 object properties when objects were shown in isolation ($Ps < 0.05$, except context and distance, $Ps > 0.5$, ns), but was not sensitive to any of these properties when objects were shown as part of a scene (all $Ps > 0.2$, ns). Visual size was correlated with the fMRI response to Objects in PPA and TOS ($Ps < 0.001$), but not RSC ($P = 0.09$), and was uncorrelated with response to Scenes in any region ($P > 0.5$, ns). These results indicate that our main finding—that scene regions are sensitive to object-based properties—will generalize to other objects beyond those that we have chosen here. [For completeness, results for LOC and EVC are included in Table 4.]

Next, we used the item data to investigate whether the 7 stimulus properties had independent or redundant effects on fMRI response. We did this by entering the item-wise fMRI responses into a step-wise regression, with each stimulus-based characteristic as a potential independent variable. In this approach, one first determines the single variable that fits the data the best and then adds or removes variables to the regression until there is no significant improvement in the fit of the model. This allows one to test whether the data can be modeled with a reduced set of covariates.

The results of this analysis indicated that a single stimulus-based characteristic sufficed to explain the variance in PPA/PHC response to Scenes (adjusted [adj.] $R^2 = 0.066$, $F_{1,198} = 14.0$, $P = 0.0002$). The most explanatory property was spatial definition ($\beta = 0.257$, $P = 0.0002$). When spatial definition was included in the model, none of the other 6 covariates explained a significant amount of additional variance ($ts < 1$, ns). When spatial definition was removed from the model, the second most explanatory property was found to be placeness ($\beta = 0.218$, $P = 0.002$): A model containing just placeness explained a significant amount of variance in PPA response (adj. $R^2 = 0.043$, $F_{1,198} = 9.9$, $P = 0.002$) with no additional variance explained by the excluded 5 covariates ($ts < 1$, ns).

For Objects, 3 characteristics were necessary to explain the variance in the PPA/PHC (adj. $R^2 = 0.043$, $F_{1,198} = 9.9$, $P = 0.002$). These were visual size ($\beta = 0.321$, $P = 0.000003$), placeness ($\beta = 0.372$, $P = 0.00008$), and physical size ($\beta = -0.198$, $P = 0.037$). The negative loading on the physical

Table 4.
Results of item-wise regression analyses in fROIs.

	Scenes (object + background)			Objects (object + no background)		
	Beta	t	p	Beta	t	P
			PPA/PHC			
Placeness	0.218	3.139	0.002**	0.267	4.040	0.000***
SpaceDefining	0.257	3.737	0.000***	0.227	3.280	0.001**
Context	0.151	2.146	0.033*	0.038	0.530	0.597
Distance	0.145	2.059	0.041*	0.073	1.031	0.304
Fixedness	0.184	2.639	0.009**	0.188	2.693	0.008**
PhysicalSize	0.193	2.761	0.006**	0.144	2.049	0.042*
VisualSize	0.009	0.125	0.901	0.325	4.837	0.000***
			RSC			
Placeness	0.205	2.945	0.004**	0.196	2.815	0.005**
SpaceDefining	0.212	3.054	0.003**	0.228	3.302	0.001**
Context	0.115	1.622	0.106	0.080	1.132	0.259
Distance	0.166	2.368	0.019*	0.174	2.491	0.014*
Fixedness	0.145	2.068	0.040*	0.222	3.205	0.002**
PhysicalSize	0.145	2.065	0.040*	0.242	3.505	0.001**
VisualSize	−0.031	−0.436	0.663	0.119	1.690	0.093
			TOS			
Placeness	0.091	1.284	0.201	0.247	3.594	0.000***
SpaceDefining	0.081	1.143	0.254	0.247	3.588	0.000***
Context	0.043	0.603	0.547	−0.018	−0.248	0.805
Distance	0.008	0.113	0.910	0.025	0.352	0.725
Fixedness	−0.013	−0.187	0.852	0.021	2.313	0.022*
PhysicalSize	0.045	0.632	0.528	0.146	2.071	0.040*
VisualSize	0.024	0.342	0.733	0.247	3.592	0.000***
			LOC			
Placeness	−0.192	−2.759	0.006**	0.139	1.977	0.049*
SpaceDefining	−0.200	−2.868	0.005**	0.058	0.813	0.417
Context	−0.126	−1.792	0.075	−0.015	−0.215	0.830
Distance	−0.181	−2.596	0.010*	−0.052	−0.727	0.468
Fixedness	−0.178	−2.544	0.012*	0.000	−0.002	0.998
PhysicalSize	−0.121	−1.708	0.089	−0.083	−1.173	0.242
VisualSize	0.028	0.396	0.692	0.016	0.225	0.822
			EVC			
Placeness	−0.065	−0.918	0.360	0.118	1.668	0.097
SpaceDefining	−0.114	−1.621	0.107	0.085	1.196	0.233
Context	−0.089	−1.261	0.209	−0.014	−0.199	0.842
Distance	−0.066	−0.929	0.354	0.011	0.153	0.878
Fixedness	−0.104	−1.476	0.142	0.126	1.794	0.074
PhysicalSize	−0.053	−0.752	0.453	0.038	0.539	0.591
VisualSize	0.022	0.309	0.758	0.373	5.654	0.000***

*** p < 0.001 ** p < 0.01 * p < 0.05

$P = 0.004$): A model containing just placeness explained a significant amount of variance in PPA response (adj. $R^2 = 0.037$, $F_{1,198} = 8.6$, $P = 0.004$) with no additional variance explained by the excluded 5 covariates ($t < 1$, ns). RSC response to objects could also be explained by a single property (adj. $R^2 = 0.054$, $F_{1,198} = 12.3$, $P = 0.001$); in this case, physical size ($\beta = 0.242$, $P = 0.001$; $t < 1.2$, ns for all other properties). When physical size was excluded, the second most explanatory property, which also sufficed to explain a significant amount of variance (adj. $R^2 = 0.047$, $F_{1,198} = 10.9$, $P = 0.001$), was spatial definition ($\beta = 0.228$, $P = 0.001$; $t < 1.4$, ns for all other properties).

For TOS, none of the stimulus properties were found to explain a significant amount of variance for Scenes, consistent with our previous results indicating that TOS was not sensitive to these properties when background was included. A model containing 3 properties sufficed to explain the variance in the response to Objects (adj. $R^2 = 0.124$, $F_{1,198} = 12.9$, $P = 0.000002$). These properties were placeness ($\beta = 0.392$, $P = 0.00005$), visual size ($\beta = 0.187$, $P = 0.006$), and distance ($\beta = -0.244$, $P = 0.01$). Of the excluded variables, spatial definition was just at the cusp of the threshold for inclusion ($P = 0.052$); the other properties were not significant ($ts < 1.4$).

These results suggest that the 6 object-based characteristics are highly redundant drivers of the PPA/PHC and RSC: Once the single best object-based characteristic is known and (in the case of PPA/PHC) the visual size of the stimulus is accounted for, additional characteristics do not explain significantly more variance in the fMRI response. Furthermore, they are consistent with the idea that some object-based characteristics are better predictors of fMRI response than others. In the PPA/PHC, the best predictors were spatial definition (for Scenes) and placeness (for Objects); in RSC, the best predictors were spatial definition (for Scenes) and physical size (for Objects). In TOS, on the other hand, the picture was more complicated: In this region, modeling the response to Objects required consideration of at least 2 object-based characteristics in addition to visual size.

Discussion

The principal finding of our study is that scene-responsive regions (PPA/PHC, RSC, and TOS) are sensitive to several object-based properties in addition to being sensitive to the categorical difference between Scenes (with-background stimuli) and Objects (without-background stimuli). We tested 6 such properties: Placeness, spatial definition, context, fixedness, physical size, and distance. Consistent with previous reports, the PPA/PHC and RSC responded more strongly to objects that were more place-related, more spatially-defined, more strongly associated with a context, more likely to be fixed in a stable location, larger in real-world size, and more distant. With some exceptions, these sensitivities were observed both when objects were shown with scenic background and also when they were shown in isolation. TOS exhibited a similar preference for objects that were more place-related, more spatially-defined, more fixed, and larger, but only when the objects were shown without scenic background. These results confirm and extend previous findings and also advance our understanding of the function of scene-responsive regions in several important ways.

size covariate was surprising in light of our previous findings that this property positively affects PPA/PHC response when considered alone. Inspection of the data revealed a small number of stimuli with large physical size that did not strongly activate the PPA. These included a cloud, the space shuttle, and the sphinx. Notably, these are all large objects that are not especially place-like, either because they lack a definite location (cloud), or because their interpretation as objects is likely to be especially salient (space shuttle, sphinx). When these 3 items were excluded from the stepwise regression, the physical size factor was no longer included in the model, leaving just one object-based characteristic plus visual size sufficient to explain PPA/PHC response to without-background stimuli.

Results in RSC were similar. As in the PPA/PHC, a single stimulus-based characteristic sufficed to explain the variance in RSC response to Scenes (adj. $R^2 = 0.040$, $F_{1,198} = 9.3$, $P = 0.003$). The most explanatory property was spatial definition ($\beta = 0.212$, $P = 0.003$). When spatial definition was included in the model, none of the other 6 covariates explained a significant amount of additional variance ($t < 1$, ns). When spatial definition was removed from the model, placeness was found to be the second most explanatory property ($\beta = 0.205$,

Most notably, the current results suggest that the 6 object properties act as redundant rather than independent drivers of fMRI response in the PPA/PHC and RSC. These properties tend to covary with each other in the real world: Larger objects tend to be more fixed, more spatially defining, viewed at further distances, more place-related, and more likely to be associated with a single context. So, it is not surprising that these quantities covaried in our naturalistic stimulus set. Indeed, a PC analysis on the item-wise ratings for these quantities plus visual size suggested that 2 factors could explain much of the variability. One of these factors loaded heavily on visual size, while the other loaded heavily on the 6 object-based characteristics. Re-analysis of the fMRI data in terms of these 2 factors suggested that they explained a considerable amount of the neural response in the PPA/PHC and RSC. Furthermore, a step-wise regression analysis of the item-wise fMRI responses found that only 1 object-based factor sufficed to explain most of the variability in the PPA/PHC and RSC response to Scenes and the RSC response to Objects, whereas 1 object-based factor plus visual size sufficed to explain most of the variability in the PPA/PHC response to Objects. These results argue in favor of redundancy rather than independence among the object-based properties.

Our data also provide insights as to which of the object-based properties best explain response in scene-selective regions. Comparison between the individual-property models indicated that the properties modulated fMRI activity to different degrees. In the PPA/PHC, the strongest modulation was observed for placeness and spatial definition, whereas in RSC the strongest modulation was observed for placeness, distance, physical size, and spatial definition. In both regions, context was the weakest modulator of fMRI response. These observations were bolstered by the results of the step-wise regression analyses, which show similar factors as most explanatory (placeness and spatial definition in the PPA/PHC, spatial definition and physical size in RSC). Given that these analyses also suggested that the various object-based properties have redundant effects, we suggest that the best way to interpret these data is to posit that some properties predict fMRI response better than others, because they are better approximations of a single underlying object-based factor that drives the PPA/PHC and RSC.

What is this true underlying factor? In the case of the PPA/PHC, it is notable that the 2 properties that best explained fMRI response were placeness and spatial definition. Objects with high placeness look like other objects, but they are treated cognitively as being indicative of a location in the world (e.g. Igloo, McDonald's Arches). Similarly, objects with high spatial definition have a size and shape that defines a local spatial coordinate frame in the same way that a scene does, even though they do not look like scenes (e.g. Goal Posts, Dentist Chair). In both cases, the PPA/PHC responds to qualities that are not immediately visible in the object that suggests that the object is spatially relevant. This spatial relevance makes the object a potential navigational landmark, and we believe that it is the interpretation of the object as a landmark or as an indicator of a place that ultimately drives object-based activity in the PPA/PHC. This account is consistent with previous findings that response in PPA/PHC (and RSC) can be increased by previously experiencing an object at a navigational decision point (Janzen and van Turennout 2004; Janzen and Jansen 2010; Schinazi and Epstein 2010) in these earlier experiments, it was the navigational history of that the object, rather than its inherent qualities, that drove the increased response. It can also explain the long-standing finding that the PPA/PHC responds more strongly to objects than to faces, since objects can occasionally be useful as landmarks, but faces are never useful because people change their locations frequently. A similar account can explain the response to object-based characteristics in RSC, but here the underlying factor seems to be more closely related to purely spatial qualities of the object such as size and distance.

In this view, the division between scene and object is something less than absolute, because an object can act as shorthand for identifying a place, or as an anchor for an implied scene. But does this mean that there is nothing special about scenes, at least as far as the PPA/PHC and RSC are concerned? Not necessarily. In the natural world, clear-cut object-like landmarks are rare, while scenes are encountered in every glance. Furthermore, scenes are places by definition, whereas objects merely have the potential to define a place. Thus, it would make sense that a system involved in spatial navigation or place recognition might be especially attuned to processing of visual and spatial features that are characteristic of scenes, for example, by creating statistical summaries of visual features over wide portions of the visual field, or by analyzing the geometry defined by extended background elements such as walls or hillsides. In this view, the ultimate outcome of the processing in the PPA/PHC, or RSC, would be the same irrespective of whether the stimulus was a scene or an object: Identification of the current location, and/or localization within a local (PPA/PHC) or global (RSC) spatial coordinate frame. But these regions might nevertheless respond more strongly to scenes than to objects (even landmark objects), because scenes contain the stimulus features to which these regions are more closely attuned, whereas objects do not. Furthermore, the extraction of spatial quantities might proceed automatically for scenes, but might require top-down assistance for isolated objects (cf. Epstein and Ward 2010).

Indeed, the fact that the PPA/PHC and RSC responded more strongly to Scene (with-background) than to Object (without-background) stimuli, over and above their response to the qualities of the objects, lends credence to the idea that these regions might have some sensitivity to visual or geometric features that are specific to scenes. Further supporting this idea are recent findings showing that the PPA/PHC processes information about the visual summary statistics of images; such information is likely to be useful for scene identification (Cant and Xu 2012). Other recent studies have shown that the PPA/PHC is sensitive to spatial frequency, another low-level visual property that tends to differ between Scenes and Objects. For example, Rajimehr et al. (2011) demonstrated a high spatial frequency bias in both human PPA/PHC and the macaque homolog, regardless of the category represented by the image. Complementing this, Zeidman et al. (2012) demonstrated an interaction between space and spatial frequency in the PPA/PHC, whereby sensitivity to spatial properties was only reliable for high spatial frequency stimuli. An interesting topic for future research would be to examine the interaction between image-based and object-based characteristics in the PPA/PHC.

Beyond the main finding that all 3 scene-selective regions (PPA/PHC, RSC, and TOS) were sensitive to object-based

properties, we observed some notable differences between these regions, which potentially speak to their specific functions. First, PPA/PHC and TOS were sensitive to the retinotopic extent of isolated objects, whereas RSC was not. This is consistent with previous accounts that PPA/PHC and TOS are more directly involved in visual analysis of stimuli, whereas RSC is more involved in mnemonic processes that are less closely linked to perception. Secondly, as previously noted, response to object-based properties was observed for both Scenes and Objects in PPA/PHC and RSC, but only for Objects in TOS. This suggests that TOS might support perceptual analysis of the spatial qualities of individual objects—the addition of the scenic background might have eliminated the ability to measure object-based effects in this region, because the spatial qualities of the focal object were diluted by the spatial qualities of the other objects in the scene. In contrast, PPA/PHC and RSC appear to analyze these quantities in a synergistic way that extracts a landmark value for the scene as a whole, which is influenced by the landmark value of the focal object but relatively unaffected by the incidental objects in the periphery. The fact that at least 2 object-based characteristics were necessary to explain the variability in the item-wise responses in TOS (in contrast to PPA/PHC and RSC, where only one characteristic sufficed) gives further credence to the idea that TOS may extract individual spatial properties, but not combine these into a single rating of landmark value. Finally, RSC responded more strongly to object-based properties when these objects were shown on a scenic background, whereas PPA/PHC responded equally strongly to object properties irrespective of whether background was present or absent. Although not directly analogous, this result echoes the findings of a recent multivoxel pattern analysis study that suggested the existence of a gradient of sensitivity across regions, whereby RSC is more sensitive to information obtained from scenes and PPA/PHC is sensitive to information obtained from both Scenes and Objects (Harel et al. forthcoming).

Finally, a novel aspect of our approach was the use of an item analysis to assess the reliability of effects across items. The results indicated that the object-based responses observed here are not an artifact of the particular stimuli used, but would generalize to a different stimulus set. Furthermore, this method allowed us to perform higher-level statistics, such as the step-wise regression, on fMRI data and enabled us to explicitly test the redundancy of these object effects. Item-based approaches have only been applied previously in a few language and social neuroimaging studies (Bedny et al. 2007; Yee et al. 2010; Dodell-Feder et al. 2011) and more recently in the visual domain (Mur et al. 2012); we suggest that they should be more widely used.

In sum, the current results confirm and extend previous reports that scene regions are sensitive to object-based properties. We found little evidence that these properties were independent; rather, they were highly redundant with each other. We suggest that these properties drive response in scene regions because they act as a cue that an object is a suitable landmark and thus indicative of a place or spatial location.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Notes

Conflict of Interest: None declared.

References

Aguirre GK, Zarahn E, D'Esposito M. 1998. The variability of human BOLD hemodynamic responses. Neuroimage. 8:360–369.

Aminoff E, Gronau N, Bar M. 2007. The parahippocampal cortex mediates spatial and nonspatial associations. Cereb Cortex. 17:1493–1503.

Amit E, Trope Y, Mehudar E, Yovel G. 2012. Activation of ventral visual cortex supports distance representation. Brain Cogn. 80:201–213.

Arcaro MJ, McMains SA, Singer BD, Kastner S. 2009. Retinotopic organization of human ventral visual cortex. J Neurosci. 29: 10638–10652.

Auger SD, Mullally SL, Maguire EA. 2012. Retrosplenial cortex codes for permanent landmarks. PLoS One. 7:e43620..

Bar M. 2004. Visual objects in context. Nat Rev Neurosci. 5:617–629.

Bar M, Aminoff E. 2003. Cortical analysis of visual context. Neuron. 38:347–358.

Bar M, Aminoff E, Schacter DL. 2008. Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. J Neurosci. 28:8539–8544.

Bedny M, Aguirre GK, Thompson-Schill SL. 2007. Item analysis in functional magnetic resonance imaging. Neuroimage. 35:1093–1102.

Cant JS, Xu Y. 2012. Object ensemble processing in human anterior-medial ventral visual cortex. J Neurosci. 32:7685–7700.

Cate AD, Goodale MA, Kohler S. 2011. The role of apparent size in building- and object-specific regions of ventral visual cortex. Brain Res. 1388:109–122.

Diana RA, Yonelinas AP, Ranganath C. 2008. High-resolution multi-voxel pattern analysis of category selectivity in the medial temporal lobes. Hippocampus. 18:536–541.

Dodell-Feder D, Koster-Hale J, Bedny M, Saxe R. 2011. fMRI item analysis in a theory of mind task. Neuroimage. 55:705–712.

Downing PE, Jiang Y, Shuman M, Kanwisher N. 2001. A cortical area selective for visual processing of the human body. Science. 293:2470–2473.

Epstein R, Harris A, Stanley D, Kanwisher N. 1999. The parahippocampal place area: recognition, navigation, or encoding? Neuron. 23:115–125.

Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. Nature. 392:598–601.

Epstein RA, Higgins JS. 2007. Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. Cereb Cortex. 17:1680–1693.

Epstein RA, MacEvoy SP. 2011. Making a scene in the brain. In: Harris L, Jenkin M, editors. Vision in 3D environments. Cambridge: Cambridge University Press.

Epstein RA, Morgan LK. 2011. Neural responses to visual scenes reveals inconsistencies between fMRI adaptation and multivoxel pattern analysis. Neuropsychologia. 50:530–543.

Epstein RA, Parker WE, Feiler AM. 2007. Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. J Neurosci. 27:6141–6149.

Epstein RA, Ward EJ. 2010. How reliable are visual context effects in the parahippocampal place area? Cereb Cortex. 20:294–303.

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. 1999. Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. Nat Neurosci. 2:568–573.

Harel A, Kravitz DJ, Baker CI. 2013. Deconstructing visual scenes in cortex: gradients of object and spatial layout information. Cereb Cortex. 23:947–957.

Henderson JM, Hollingworth A. 1999. High-level scene perception. Ann Rev Psychol. 50:243–271.

Janzen G, Jansen C. 2010. A neural way finding mechanism adjusts for ambiguous landmark information. Neuroimage. 52:364–370.

Janzen G, van Turennout M. 2004. Selective neural representation of objects relevant for navigation. Nat Neurosci. 7:673–677.

Kanwisher N. 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. Proc Natl Acad Sci U S A. 107:11163–11170.

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 17:4302–4311.

Konkle T, Oliva A. 2011. Organizing visual object knowledge by real-world size in ventral visual cortex. J Vis. 11:883.

Kravitz DJ, Peng CS, Baker CI. 2011. Real-world scene representations in high-level visual cortex: it's the spaces more than the places. J Neurosci. 31:7322–7333.

MacEvoy SP, Epstein RA. 2011. Constructing scenes from objects in human occipitotemporal cortex. Nat Neurosci. 14:1323–1329.

Morgan LK, Macevoy SP, Aguirre GK, Epstein RA. 2011. Distances between real-world locations are represented in the human hippocampus. J Neurosci. 31:1238–1245.

Mullally SL, Maguire EA. 2011. A new role for the parahippocampal cortex in representing space. J Neurosci. 31:7441–7449.

Mur M, Ruff DA, Bodurka J, De Weerd P, Bandettini PA, Kriegeskorte N. 2012. Categorical, yet graded—single-image activation profiles of human category-selective cortical regions. J Neurosci. 32:8649–8662.

Park S, Brady TF, Greene MR, Oliva A. 2011. Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. J Neurosci. 31:1333–1340.

Park S, Intraub H, Yi DJ, Widders D, Chun MM. 2007. Beyond the edges of a view: boundary extension in human scene-selective visual cortex. Neuron. 54:335–342.

Pruessner JC, Kohler S, Crane J, Pruessner M, Lord C, Byrne A, Kabani N, Collins DL, Evans AC. 2002. Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. Cereb Cortex. 12:1342–1353.

Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RB. 2011. The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. PLoS Biol. 9: e1000608.

Ross MG, Oliva A. 2010. Estimating perception of scene layout properties from global image features. J Vis. 10:21–25.

Schinazi VR, Epstein RA. 2010. Neural correlates of real-world route learning. Neuroimage. 53:725–735.

Spiridon M, Kanwisher N. 2002. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron. 35:1157–1165.

Walther DB, Caddigan E, Fei-Fei L, Beck DM. 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. J Neurosci. 29:10573–10581.

Walther DB, Chai B, Caddigan E, Beck DM, Fei-Fei L. 2011. Simple line drawings suffice for functional MRI decoding of natural scene categories. Proc Natl Acad Sci U S A. 108: 9661–9666.

Ward EJ, MacEvoy SP, Epstein RA. 2010. Eye-centered encoding of visual space in scene-selective regions. J Vis. 10:6.

Yee E, Drucker DM, Thompson-Schill SL. 2010. fMRI-adaptation evidence of overlapping neural representations for objects related in function or manipulation. Neuroimage. 50:753–763.

Zeidman P, Mullally SL, Schwarzkopf DS, Maguire EA. 2012. Exploring the parahippocampal cortex response to high and low spatial frequency spaces. Neuroreport. 23:503–507.